

BMJ Open Models predicting survival to guide treatment decision-making in newly diagnosed primary non-metastatic prostate cancer: a systematic review

David Thurtle,¹ Sabrina H Rossi,^{1,2} Brendan Berry,¹ Paul Pharoah,³ Vincent J Gnanaprasam^{1,2}

To cite: Thurtle D, Rossi SH, Berry B, *et al*. Models predicting survival to guide treatment decision-making in newly diagnosed primary non-metastatic prostate cancer: a systematic review. *BMJ Open* 2019;**9**:e029149. doi:10.1136/bmjopen-2019-029149

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-029149>).

Received 14 January 2019
Revised 15 May 2019
Accepted 22 May 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Surgery, University of Cambridge, Cambridge, UK

²Department of Urology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

³Cancer Epidemiology, University of Cambridge, Cambridge, UK

Correspondence to

David Thurtle;
dt433@cam.ac.uk

ABSTRACT

Objectives Men diagnosed with non-metastatic prostate cancer require standardised and robust long-term prognostic information to help them decide on management. Most currently-used tools use short-term and surrogate outcomes. We explored the evidence base in the literature on available pre-treatment, prognostic models built around long-term survival and assess the accuracy, generalisability and clinical availability of these models.

Design Systematic literature review, pre-specified and registered on PROSPERO (CRD42018086394).

Data sources MEDLINE, Embase and The Cochrane Library were searched from January 2000 through February 2018, using previously-tested search terms.

Eligibility criteria Inclusion required a multivariable model prognostic model for non-metastatic prostate cancer, using long-term survival data (defined as ≥5 years), which was not treatment-specific and usable at the point of diagnosis.

Data extraction and synthesis Title, abstract and full-text screening were sequentially performed by three reviewers. Data extraction was performed for items in the CHECKlist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies checklist. Individual studies were assessed using the new Prediction model Risk Of Bias ASsessment Tool.

Results Database searches yielded 6581 studies after deduplication. Twelve studies were included in the final review. Nine were model development studies using data from over 231 888 men. However, only six of the nine studies included any conservatively managed cases and only three of the nine included treatment as a predictor variable. Every included study had at least one parameter for which there was high risk of bias, with failure to report accuracy, and inadequate reporting of missing data common failings. Three external validation studies were included, reporting two available models: The University of California San Francisco (UCSF) Cancer of the Prostate Risk Assessment score and the Cambridge Prognostic Groups. Neither included treatment effect, and both had potential flaws in design, but represent the most robust and usable prognostic models currently available.

Conclusion Few long-term prognostic models exist to inform decision-making at diagnosis of non-metastatic prostate cancer. Improved models are required to inform

Strengths and limitations of this study

- Comprehensive and focused search strategy on prognostic models built around survival outcomes rather than shorter-term surrogates.
- Contemporary clinical review of model accuracy and usability, to inform clinical practice.
- Thorough bias assessments of individual studies, utilising the newly developed PROBAST tool.
- Unable to assess bias across studies, publication bias and selective reporting.

management and avoid undertreatment and overtreatment of non-metastatic prostate cancer.

INTRODUCTION

Prostate cancer (PCa) is the most common male cancer and its incidence is increasing, with 1.3 million estimated new cases globally in 2018.^{1,2} The vast majority of new presentations in the UK (>80%) are with localised or locally advanced disease representing a significant healthcare and economic burden.³ Treatment decisions, in this growing group of men, are notoriously complex with the risk of progression and psychological impact of a cancer diagnosis balanced against significant potential morbidity associated with treatment. These latter problems can be very significant with rates of erectile dysfunction as high as 79% and 66% 3 years after prostatectomy and radiotherapy, respectively, and incontinence rates of 20% and 3%, respectively.⁴ As a result of this, the uptake of conservative management is increasing with rising confidence in using active surveillance.^{5,6} The predominant decision dilemma is therefore at the point of diagnosis when men are considering treatment options and indeed whether treatment is needed, particularly given the fact that many men may have indolent disease and are

more likely to die of non prostate cancer related causes. This point is highlighted in the work of Lloyd *et al* who showed that the lifetime incidence of being diagnosed with the disease is 13.4% but the risk of dying from the disease is only 4.3%.⁷

Most national guidelines currently risk-stratify men according to modified versions of the three-stratum D'Amico classification system, first proposed in 1998⁸. This used biochemical recurrence as the primary outcome from a cohort of men all managed by radical treatment. However, biochemical recurrence is known to be a poor surrogate for survival and many men will no longer undergo radical treatment.^{5 6 9} The value of this system is therefore questionable, especially given its use has moved from predicting radical therapy outcomes to counselling men at diagnosis about whether to have surveillance or treatment. Alternative risk models have been proposed to delineate smaller groups using standard variables (prostate-specific antigen (PSA), Gleason score, T-stage) or which integrate additional parameters, such as biopsy characteristics.^{10–12} However, many are built around single-centre data, using PSA-screened and heavily radically-treated populations, making them less applicable to the fundamental decision dilemma of whether treatment is needed in the first place.

Prognostic models, according to the American Joint Committee on Cancer (AJCC), should use survival itself as an endpoint, which is less equivocal and more robust.¹³ Using survival is especially important in prostate cancer given the slow natural history of the disease. In other tumour-types, high quality prognostic models using long-term survival are already integrated into routine practice and endorsed by the AJCC.¹⁴ However, no prognostic model for PCa has yet been endorsed, nor to our knowledge, is any such model widely-used in routine clinical practice. Models integrating the impact of radical treatment compared with conservative management would be particularly powerful.

Our objective in this study was therefore to perform a rigorous systematic search of the literature to identify available prognostic models built specifically around long-term patient survival, available for use at the point of a new diagnosis of primary non-metastatic PCa. Our aims were to establish (i) what models were available (ii) their accuracy in terms of discrimination and calibration and (iii) their generalisability, external validation and clinical utility.

METHODS

The study protocol followed the recommendations of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.¹⁵ The review aim, search strategy and study inclusion and exclusion criteria were framed using the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS).¹⁶ The search strategy was informed by previous similar studies, including

Box 1 Eligibility criteria for study inclusion and exclusion

Study inclusion criteria

All of the following inclusion criteria must be met:

- ▶ Studies reporting models based on men with non-metastatic prostate cancer.
- ▶ Studies evaluating 'long-term' (≥5 years) cancer-specific or overall survival outcomes.
- ▶ Studies reporting models in screened or non-screened populations.
- ▶ Studies including men undergoing more than one treatment option.
- ▶ Models available for use at the point of diagnosis – that is, pre-treatment.
- ▶ The model includes more than one parameter, that is, multivariable.

Study exclusion criteria

Any of the following is a reason to exclude a study:

- ▶ Any article that is not an original study (eg, reviews, commentary, editorials, corrigendums, letters).
- ▶ Conference proceeding or abstract from poster/oral communication only.
- ▶ Study where data cannot be derived to contribute to a primary or secondary outcome of this systematic review.
- ▶ Studies pertaining only to men with advanced/metastatic disease.
- ▶ Studies pertaining exclusively to men **after** an active treatment option for example, after radical prostatectomy.
- ▶ Studies of single biomarkers or single parameters only.
- ▶ Studies including men exclusively undergoing a single treatment type.

publications which tested and recommended search terms for risk-prediction models.^{17 18} The full systematic review protocol was pre-specified and registered through PROSPERO, reference CRD42018086394 (<https://www.crd.york.ac.uk/prospero/>) and is available in the online supplementary files.

In summary, this was a review of studies reporting multi-variable long-term survival models for use at the point of diagnosis for men newly diagnosed with non-metastatic PCa. Long-term survival, was defined as ≥5 years following diagnosis. We focused on publications subsequent to January 2000 to increase relevance to modern practice. For inclusion, studies needed to include men undergoing more than one treatment type and models should be multivariable. Models for cancer-specific or all-cause survival outcomes were potentially eligible, and any eventual model types were allowable. Both model development and model validation studies were eligible. Single-parameter or single-treatment studies were excluded. Comprehensive study inclusion and exclusion criteria are shown in **box 2**.

Information sources

Studies were identified by searching MEDLINE, Embase and Cochrane Library from 1 January 2000 to 28 February 2018. Detailed search strategies for each database are available in the online supplementary data. Highly relevant but excluded articles were recorded, collated and the references analysed for additional studies.

Data management

Search results were exported into Covidence software, an online screening platform endorsed by Cochrane (covidence.org). Title and abstract screening and full-text screening were sequentially performed by a team of three reviewers. Prior to screening, a pilot screening process was conducted for calibration of screening between reviewers. Reviewers were not blinded to study authors, institution, publication journal or year of publication.

Data items

The full list of data items extracted from each included study is recorded in the protocol. These were informed by the CHARMS checklist¹⁶ and included: (1) Study design, (2) Characteristics of study participants, (3) Outcomes, (4) Candidate predictors, (5) Sample size, (6) Missing data, (7) Statistical methods, (8) Model performance and evaluation and (9) Usability. Model performance, assessed by discrimination was the principal summary measure.

Bias assessment

To assess the validity of eligible studies, individual studies were assessed for bias using the new Prediction model Risk Of Bias ASsessment Tool (PROBAST).¹⁹ The PROBAST tool assesses both risk of bias and applicability of both model development studies and validation studies.²⁰

Patient and public involvement

There was no formalised patient and public involvement in the design or conduct of this review. Preliminary results were presented to a departmental patient and public involvement group, and their comments used to inform the write-up of the review.

RESULTS

Study selection

The search of Cochrane, MEDLINE and Embase yielded 6581 studies after deduplication. Sixteen additional studies were identified by reviewing the references of excluded but relevant studies. A total of 12 studies were eligible for inclusion in the final review.^{21–32} Two of these had not been summarised in previous reviews.^{26 27} The PRISMA flow diagram is shown in figure 1, including the reasons for exclusion at full-text screening. Only one reason for exclusion was assigned to each study, when multiple reasons may have been present. Nine of the final 12 included studies were model development studies, three were model validation studies. Two of these external validations were of models already included as model development studies.^{30 31} One study related to an external validation of the Cancer of the Prostate Risk Assessment (CAPRA) score against mortality.³² The original CAPRA model development study however did not meet the eligibility criteria as it was developed against recurrence rather than long-term survival.¹¹

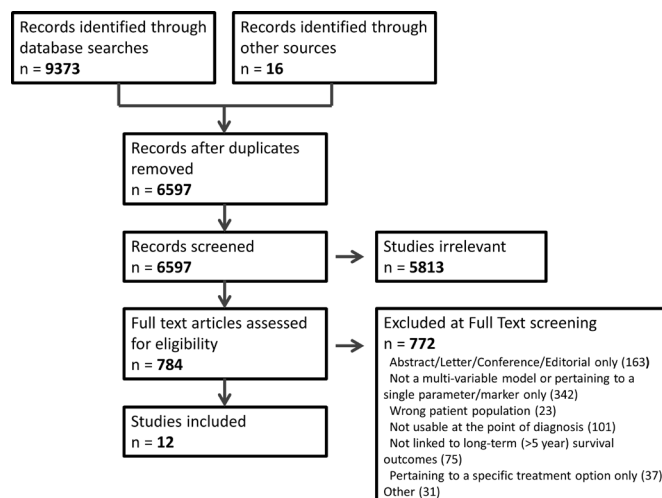


Figure 1 PRISMA flow diagram depicting the flow of information through the different phases of the review. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Study characteristics

Characteristics of the included studies' participants and settings are summarised in table 1. The model development studies used data from over 231 888 men. However, two studies used analytical cohorts from the same registry (Surveillance, Epidemiology and End Results (SEER)).^{23 29} Two studies used data from single US centres, two used data from groups of four hospitals, and five were from regional or national multicentre registries. Eligibility criteria of patients into studies varied significantly, with some models using very specific or selected sub-cohorts only. For example Nguyen *et al* included only consenting men, undergoing radiotherapy (RT) or radical prostatectomy (RP), with at least one intermediate or high-risk feature but \leq T3b disease.²² The treatment cohorts included, and whether treatment effect was a parameter in the final model was highly variable. Only six of the nine models included any men who had been managed conservatively (including 'watchful waiting' or 'conservative management'); none of the models described or defined a specific 'active surveillance' cohort.^{21 23 25 27–29} Only three of the nine models included treatment as a predictor variable - none of which were externally validated.^{24 25 28} Median follow-up was relatively short within all model development cohorts, with the longest reported follow-up being 7.6 years.²⁶

Results of individual studies

The final results of individual studies are summarised in table 2. The primary outcome was cancer-specific mortality only in one study,²⁷ and was overall-survival only in another study.²⁵ The remainder reported measures of both. The study designs, model-types and performance metrics varied markedly. We therefore focused on describing the studies, results, applicability and availability. Significant heterogeneity in the question being asked and statistical methods of validation between

Table 1 Study characteristics of model development studies – participants and setting

Study	Source of data	Country	Years of diagnosis	Eligibility summary	Overall (n)	Number of PCa deaths	Number of overall deaths	Treatment cohorts included	Median follow-up
Margel <i>et al</i> 2014 ²¹	Registry (Ontario Cancer Registry)	Canada	1994–2008	Aged 66+, with diabetes prior to PCa diagnosis and pathology reports available	4001	321	1395	Surgery, RT, WW, ADT	4.7 years
Nguyen <i>et al</i> 2009 ²²	Hospital data (four centres)	USA	1965–2002	Consenting men who had RP or RT or RT+ADT for up to T3b PCa at four centres with at least one int/high risk factor (PSA>10, GL7+, t2b+, PSAV>2.0ng/mL/yr)	1063	178	NR	Surgery, RT, RT+ADT	5.6 years
Feuer <i>et al</i> 2012 ²³	Registry (SEER)	USA	1995–2005	PCa, aged 40–96, with complete information and cause of death (and linked medicare data if aged 66+)	203 546	NR	NR	All	NR
Kutikov <i>et al</i> 2012 ²⁴	Registry (CaPSURE)	USA	NR	Localised PCa, undergoing RT or RP	6091	167	983	Surgery, RT	4.4 years
Cowen <i>et al</i> 2006 ²⁵	Hospital data (single centre)	USA	1987–1989	Clinically localised PCa (<cT3) with known initial treatment strategy	506	NR	321	Surgery, RT, other (inc CM, cryo)	NR
Kerkmeijer <i>et al</i> 2016 ²⁶	Hospital data (four centres)	Belgium & Holland	1989–2008	Treated PCa patients only	3383	149	628	Surgery, RT, brachytherapy	7.6 years
Gnanapragasam <i>et al</i> 2016 ²⁷	Registry (NCRS Eastern)	UK	2000–2010	All clinically localised PCa with intact data for included variables	10 139	462	1557	Surgery, RT, CM, brachytherapy	6.9 years
Tewari <i>et al</i> 2004 ²⁸	Hospital data (single centre)	USA	1990–1997	Men under 75 with localised PCa, negative bone scan and intact data including zip code	3159	NR	NR	Surgery, RT, CM	6.2 years
Howlader <i>et al</i> 2014 ²⁹	Registry (SEER)	USA	1992–2009	All PCa patients (the model including comorbidity includes men aged 66+ only)	NR	NR	NR	All	NR

ADT, androgen deprivation therapy; CaPSURE, UCSF Cancer of the Prostate Strategic Urologic Research Endeavor; CM, conservative management; inc, including; int, intermediate; NR, not recorded; PCa, prostate cancer; PSA, prostate-specific antigen; PSAV, prostate-specific antigen velocity; RP, radical prostatectomy; RT, radiotherapy; SEER, Surveillance, Epidemiology and End Results; WW, watchful waiting.

Table 2 Study characteristics of model development studies: analysis and model output

Study	Outcome(s) or endpoint(s)	Time frame of predictions	Model type	Statistical methods	Included predictors	Missing data handling	Number of groups	Internal validation method	Discrimination results (c-index or AUC)	Calibration results	Model availability
Margel <i>et al</i> 2014 ²¹	PCSM and ACM	5 year	N/A	Nested Cox model	Two models: 1=age, Johns Hopkins comorbidity group, year of cohort entry, socioeconomic status and housing status. 2=above + Gleason grade and cancer volume.	Complete case analysis	N/A	Bootstrapping	Admin data only model ACM c-index 0.70 (95% CI 0.69 to 0.71) PCSM 0.76 (95% CI 0.74 to 0.78) Extended model ACM 0.74 (95% CI 0.73 to 0.75) PCSM: 0.85 (95% CI 0.83 to 0.87)	NR	No
Nguyen <i>et al</i> 2009 ²²	PCSM and ACM	5 year	Group-based stratification	Fine & Grey	PSA, PSAV, stage, Gleason	NR	four groups	KM curves only	NR	NR	Yes - paper
Feuer <i>et al</i> 2012 ²³	PCSM, OCM and LE	3–5 and 10 year	Nomogram	Cox model	Preclinical model: stage, Gleason, age, race, marital status, (comorbidity if 66+)	Complete case analysis	'Individualised'	Time dependent AUC	10-year AUCs: Pretreatment model: <66 years, PCSM 0.82, OCM 0.69, 66+ yrs PCSM 0.87, OCM 0.61	Plots	No
Kutikov <i>et al</i> 2012 ²⁴	PCSM and NPCM	10 year	Nomogram	Fine & Grey	Age, race, comorbidity, primary treatment (RT vs RP), receipt of ADT and modified CAPRA score	Imputed mean values	Nomogram	NR	NR	NR	Yes - paper
Cowen <i>et al</i> 2006 ²⁵	Overall survival	5, 10 and 15 year and median LE	Nomogram	Cox model	Age, Charlson, performance status, angina, BMI, smoking, marital status, PSA, Gleason, treatment	Imputation	260 points	Within complete case sub-cohort	c-index 0.73	Plots	Yes - paper
Kerkmeijer <i>et al</i> 2016 ²⁶	DSS and OS	10 year DSS	Nomogram	Cox model	T-stage, grade (on 1–3 scale) PSA, age	Single imputation	16 points - five groups	Bootstrapping	c-index 0.78 (95% CI 0.74 to 0.82) DSS, 0.68 (95% CI 0.65 to 0.70) OS	Plots	Yes - paper
Gnanapragasam <i>et al</i> 2016 ²⁷	PCSM	N/A	Group-based stratification	Cox model	PSA, grade group, T-stage integers	Complete case analysis	five groups	Random validation cohort	c-index 0.75 (95% CI 0.72 to 0.77) for PCSM	Plots	Yes - paper and online model
Tewari <i>et al</i> 2004 ²⁸	PCSM and ACM	10 year	Look up tables	Cox model	Age, grade (three groups), PSA, treatment, comorbidity, race	Complete case analysis	n/a	Unclear	0.63 PCSM, 0.69 OS	NR	Yes - paper
Howlader <i>et al</i> 2014 ²⁹	OS, PCa death, other cause death	5 year	Look up tables	Unclear	Age, stage (localised/regional/distant), comorbidity	NR	n/a	NR	NR	NR	Yes - paper

ACM, all cause mortality; AUC, area under the curve; BMI, body mass index; CAPRA, Cancer of the Prostate Risk Assessment; DSS, disease specific survival; KM, Kaplan-Meier; LE, life expectancy; N/A, not available; NPCM, non prostate cancer mortality; NR, not recorded; OCM, other cause mortality; OS, overall survival; PCa, prostate cancer; PCSM, prostate cancer specific mortality; PSA, prostate-specific antigen; PSAV, prostate-specific antigen velocity.

studies meant that attempts to meta-analyse data would not be appropriate. Modelling techniques varied, with the majority of studies using a proportional hazards model (Cox or Fine and Grey) although none reported assessing whether the proportional hazards assumption was valid. Included studies did not report any flexible parametric approaches to deal with continuous variables, and the majority used group categorisations of these variables. Reporting of model accuracy was inconsistent. Considering area under the curve and c-indices synonymously, six of the nine studies reported some measure of discrimination with values ranging from 0.63 to 0.90 for PCa survival outcomes, although this higher figure was derived within a small elderly sub-cohort^{23 28} and 0.58 to 0.73 for overall outcomes.^{23 25} Only four of the nine studies reported assessing calibration in some capacity.^{23 25–27} Relative performance within particular sub-populations were not generally reported.

Validation

Seven of the nine model development studies reported internal validation. Two reported using bootstrapping, and one used a separate 40% random sample of the original dataset for internal validation.^{21 26 27} An additional three external validation papers were included (table 3). Of note, each of these included the author of the original model within their author list, suggesting these are not completely independent validations. These three models each used large numbers of subjects over comparable timeframes to their model development study. Discrimination of the Cambridge Prognostic Groups, SEER Cancer Survival Calculator and CAPRA scores were comparable for prostate cancer specific mortality (PCSM) at 0.81, 0.81 and 0.80, respectively, over 5 years^{30–32} (table 4). Discrimination was poorer for overall mortality at 0.71 in the latter study. These external validation papers performed quite well on individual bias assessment (table 5). In the CAPRA validation paper estimates were reported to be ‘adjusted’ for age and treatment type, such that it is unclear whether the reported accuracy reflects that of the usable model.³²

Risk of bias

Risk of bias within studies is summarised in both table 5 and figure 2. Frequent concerns were observed with respect to participant selection and inclusion, particularly with respect to reporting or allowing for missing data. The outcome of death was well-defined and unambiguous in the majority of studies. Every included study had at least one parameter for which there was high concern of bias – leading to a high overall judgement of bias in the PROBAST tool. Concerns about applicability to the review question were present in more than half of the studies.

Usability

All of the included studies reported models that were nomograms, look-up tables or grouping stratifications. Seven of the nine studies were clinically usable through

the publication itself. One was also available in a dedicated website.³¹ The SEER Cancer Survival Calculator was never launched online, and the publication itself does not provide sufficient detail to use the model.²³ The model by Margel *et al* was available but not usable as it included year of entry as a predictor variable, making the model usable only on retrospective series rather than in future individual cases. Indeed, this model was developed to answer the question of whether pathological information adds to a prognostic model, rather than being intended for use at diagnosis.²¹

DISCUSSION

Principal findings

Treatment decisions at the point of diagnosis of non-metastatic PCa should be informed by the likely prognosis of the disease. Despite finding a number of published prognostic models, there remains a lack of well-validated, unbiased, generalisable models for use at diagnosis. In particular, there was a lack of external validation and dearth of models that compare outcomes between conservative management and radical treatment.

Prior evidence

A number of previous reviews have assessed ‘prediction’ or ‘risk’ models in their broadest sense.^{33–35} Shariat *et al* have previously published a thorough catalogue of available predictive models in PCa – predicting everything from detecting PCa in the initial biopsy setting to survival. Other reviews have focused exclusively on outcomes following radical prostatectomy or radiotherapy.^{36 37} Lughezzani *et al* in 2010 for example summarised models predicting Gleason score upgrading, pathological stage, life expectancy, perioperative mortality, postoperative biochemical recurrence or functional outcomes in addition to PCSM after prostatectomy.³⁶ The paucity of models using mortality as an outcome was particularly noted and the study concluded that no tools were capable of quantifying the benefit of RP relative to other treatment modalities.³⁶ In 2009, a separate review by Shariat *et al* summarised available models into eight groups. One of these groups was of models predicting survival. This group included only four models, all of which related to advanced or metastatic disease requiring hormonal therapy, or androgen-insensitive disease.³³ Green *et al* explored more historical literature from 1966 until 2012.³⁵ They included four studies which looked at life expectancy in men with localised PCa, two of which were included in our study.^{25 28} The other two were models from Albertsen *et al* in 1996 which was prior to our study dates, and from Walz *et al* in 2007 which focused exclusively on non-cancer mortality and therefore did not meet our eligibility criteria.^{38 39} A recent review into decision-making tools again took an overview of tools available for use at different points in the patient pathway.⁴⁰ Here, the only mentioned models that predicted survival were all post-prostatectomy models.⁴⁰ In 2015 Kent and

Table 3 Study characteristics of external validation studies: participants and setting

Study	Source of data	Country	Years of diagnosis	Eligibility summary	Overall (n)	Number of PCa deaths	Number of overall deaths	Treatment cohorts included	Median follow-up
Gnanapragasam <i>et al</i> 2018 ³¹	Registry (PCBaSe) & hospital data (single Singapore centre)	Sweden & Singapore	2000–2010	All non-metastatic PCa with intact data on included variables	Sweden: 72 337 Singapore: 2550	7162 (Sweden) 142 (Singapore)	23 083 (Sweden) 408 (Singapore)	All	7 years Sweden, 4.1 years Singapore
Feuer <i>et al</i> 2014 ³⁰	Hospital data (single US centre)	USA	2001–2008	Aged 40–94, intact staging details & 1 year of information prior to diagnosis	1102	NR	NR	All	NR
Cooperberg <i>et al</i> 2009 ³²	Registry (CaPSURE)	USA	<July 2007	Localised PCa (<T3b), with >6/12 follow-up, treatment known and intact data	10 627	NR	1833	All	5.9 years

NR, not recorded; PCa, prostate cancer; PCBaSe, Prostate Cancer dataBase of Sweden.

Vickers reported on ‘gross deficiencies’ in current tools for prediction of non-PCa death and concluded that they were unable to identify a suitable life expectancy tool.⁴¹ These previous reviews all suggest there is a lack of focus on and availability of good quality survival models.

Interpretation of findings

Although a number of previous reviews have been published, ours represents the most systematic and contemporary work focused towards the decision dilemma that patients and clinicians face, as it includes only models that are available at the pretreatment stage and that are not treatment-specific. This review demonstrates that only a small number of models for this setting have been published using long-term survival as an endpoint, and only three have been externally validated. However, the number of events reported in these three validation studies were well in excess of the 100 suggested as the minimum number needed for adequate validation.⁴² Within the external validations, model discrimination of up to 0.81 was reported for disease related mortality.^{30 31} The included studies highlight the potential for using large datasets to develop prognostic models. These have the advantage of providing data from ‘real-world’ settings, outside of the clinical trial context or data exclusively from specialised centres. Our included studies commonly used elements of good study design that would be in keeping with the AJCC acceptance criteria for risk models.¹³ For example, criteria that were seen in all studies, were that the prognostic time-zero was well-defined, and that model developments have been published in well regarded peer-reviewed journals. However, other criteria were lacking such as reporting measures of discrimination, assessing calibration and thorough reporting of missing data in validation work.

The most robust tools we found are the three that have been externally validated – namely the Cambridge Prognostic Groups, The SEER Cancer Survival Calculator and the UCSF CAPRA score. Of note, the Cambridge Prognostic Group stratification criteria has not previously been summarised in prior reviews, as both its development and validation have been published in the last few years.^{27 31} However, the SEER calculator has never been released for public use. Each of the other two models have significant shortcomings by disregarding treatment effect, focusing only on disease-specific mortality and ignoring comorbidity such that their value at the point of diagnosis is diminished. Importantly, each of the three external validation studies also had potential flaws in their design, using predominantly complete-case analyses and non-independent authors.^{30–32}

Many of the included studies used historical cohorts. This is a necessity when using long-term survival as an outcome, but, raises issues of generalisability to men diagnosed in the contemporary setting. The uptake of prebiopsy multiparametric MRI and targeted biopsies, may for instance impact on the type of PCa detected and affect generalisability of previous models to current practice.⁴³

Table 4 Study characteristics of external validation studies: analysis and model output

Study	Outcome(s) or endpoint(s)	Time frame of predictions	Model type	Statistical methods	Included predictors	Missing data handling	Number of groups	Discrimination results	Calibration results	Model availability
Gnanapragasam <i>et al</i> 2018 ³¹	PCSM	10 year mortalities	Five groups	N/A	PSA, grade group, T-stage integers	Complete case analysis	Five groups	c-index 0.81 for PCSM in Sweden, 0.79 in Singapore	NR	Yes - papers and online
Feuer <i>et al</i> 2014 ³⁰	DSS and OS	1, 3, 5, 7, 10 years	Nomogram - online	As per development study ³⁰	Grade (three categories), stage, marital status, age, race, year of diagnosis and comorbidity score	Complete cases + weighted averages for marital status	'Individualised'	AUCs (5 year): 0.81 PCSM, 0.81 OCM, (10 year): 0.77 PCSM 0.76 OCM	Curves	No
Cooperberg <i>et al</i> 2009 ³²	PCSM and ACM	Five and 10 year survival	Score	Cox prop hazards	CAPRA (& primary treatment, age and comorbidity in multivariable analysis)	Mixed	10 groups	PCSM 0.80 OCM 0.71	NR	Yes - papers and online

ACM, all cause mortality; AUC, area under the curve; CAPRA, Cancer of the Prostate Risk Assessment; DSS, disease specific survival; N/A, not available; NR, not recorded; OCM, other cause mortality; OS, overall survival; PCSM, prostate cancer specific mortality; PSA, prostate-specific antigen.

Another issue was of small cohorts and low numbers of events in some cases, often as a result of relatively short follow-up. The value of 5 year outcomes themselves is questionable given randomised control trial data on survival in non-metastatic PCa would suggest cancer-survival over this timeframe is incredibly high.⁴⁴ An important feature of any prognostic model should be its usability and applicability to a man diagnosed today. However, with unquantified missing data this applicability becomes less clear. Dealing with PCSM in isolation may also be problematic, not least because of the importance of competing risks of death in a disease that affects older men.⁴⁵ In this review we see that all but two included models were derived from North American data, such that generalisability to European or UK men could be questioned, with differing approaches to PSA-testing and different healthcare contexts. This is particularly relevant with regards to screening, whereby historical American cohorts are likely to have been detected through PSA screening, such that issues of lead time bias and detection of cancers that would not be considered clinically-significant may affect model generalisability to modern practice.

Access for clinicians and patients to know about and use models should be easy with modern web-based software. However, we found models were often only available in paper nomograms published with the article. Rather than online resources increasing the availability of models, there is the suggestion of the opposite occurring, with sites such as www.nomogram.org and www.clinicriskcalculators.org which were previously cited in reviews no longer being available online.³⁶ Indeed, the SEER Cancer Survival Calculator, one of the most promising models we explored, which was also externally validated, was never made available online or for clinical use.^{23 30} The UCSF CAPRA score (www.urology.ucsf.edu) and Cambridge Prognostic Groups (www.cambridgeprognosticgroup.com) on the other hand are freely available online. Online publication does not in itself difficulty in accessing models, or a reluctance to fully share coefficients may partly explain the lack of external validations by researchers outside of the original models' authorship group. Another hurdle to acceptance will be the 'face validity' of a model. For example, in any PCa prognostic model clinicians would likely expect grade, PSA and stage to be incorporated as a minimum set of variables; each of which has been shown to be independently prognostic.⁴⁶⁻⁴⁸ However, two of the models based on the largest dataset failed to include PSA which is inadequately recorded in the SEER database.^{23 30} A number of the other models used 3-strata grade classifications, rather than the full Gleason grade system, or the contemporary grade group system, which would again seem inadequate for a modern PCa model.⁴⁹

A key step in the development of any prognostic model, following validation, should be a clinical impact study to quantify whether the model's use improves decision-making and patient outcomes within a comparative design.⁵⁰ Reviewing clinical impact studies was beyond

Table 5 Individual study bias and applicability summary using the PROBAST tool

Study	Risk of bias			Concern regarding applicability			Overall		
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	Risk of bias	Applicability
Model development studies									
Margel et al 2014 ²¹	high	high	low	high	high	high	low	high	high
Nguyen et al 2009 ²²	high	high	high	high	high	low	high	high	high
Feuer et al 2012 ²³	high	unclear	low	high	high	low	low	high	high
Kutikov et al 2012 ²⁴	unclear	low	low	high	low	high	low	high	high
Cowen et al 2006 ²⁵	high	unclear	unclear	high	unclear	low	low	high	unclear
Kerkmeijer et al 2016 ²⁶	low	low	high	unclear	low	low	low	high	low
Gnanapragasam et al 2016 ²⁷	high	low	low	high	low	low	low	high	low
Tewari et al 2004 ²⁸	high	low	low	high	unclear	low	low	high	unclear
Howlader et al 2014 ²⁹	unclear	high	low	high	high	low	low	high	high
External validation studies									
Gnanapragasam et al 2018 ³¹	high	low	low	high	low	low	low	high	low
Feuer et al 2014 ³⁰	unclear	low	unclear	unclear	low	low	low	unclear	low
Cooperberg et al 2009 ³²	low	high	low	high	low	low	low	high	high

PROBAST, Prediction model Risk Of Bias Assessment Tool.

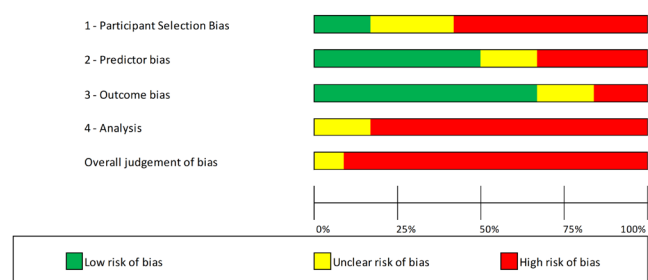


Figure 2 Bias assessment summary (across both the model development and model validation studies).

the scope of this formalised review, but no such studies were found on simple literature review. These studies not only assess whether use of the model is an improvement on standard care, but also enable the study of factors that may affect implementation into care, such as the acceptability and ease of use to clinicians or patients, which can be difficult to assess in a review such as this.⁵⁰ Impact studies can also help to bridge the gap between clinical validity and clinical utility, as utility of a model is not proportional to its prognostic capabilities. This has recently been explored further in a review on the UCSF CAPRA score, which confirmed its prognostic capacities, but was unable to demonstrate clinical usefulness – particularly when deliberating between different treatment strategies.⁵¹

Strengths and limitations at review level

While this review has particular strengths in its broad coverage of the literature and search strategy, we recognise potential limitations. Although we have assessed bias within individual studies we recognise that risk of bias will also exist across studies, driven particularly by publication bias and selective reporting within studies which we were unable to assess. Other limitations to our review may relate to our timeframe of inclusion of studies only from 2000 onwards. Models developed prior to this time may have undergone more thorough testing or validation and clinical impact assessment. However, our rationale for focusing on this contemporary time period was to investigate models appropriate to modern management; with significant changes having taken place in patient management and diagnostic practice since that time.^{43 49}

We recognise that exciting developments are also underway to propose genomic or biomarker-based prognostic indicators.⁵² Many of these are currently reported as single parameter studies, rather than being incorporated into existing models. As such these would not meet the eligibility criteria for this review. As others have suggested, any incremental value of these models should be assessed against ‘a gold-standard multivariable clinical prognostic model’.⁵³

CONCLUSION

Very few long-term prognostic models exist to inform the predominant decision dilemma of whether to undergo treatment or not after first diagnosis of non-metastatic prostate cancer. Current models are limited by inadequate external validation and fall short of many of the expectations of an unbiased, high-quality prognostic model.¹³ The most robust available tools are the Cambridge Prognostic Groups and the UCSF CAPRA score. However, both have significant shortcomings and are limited in their applicability at diagnosis by failing to include treatment effect, and disregarding non-cancer mortality. Work should focus on developing prognostic models built on long-term survival outcomes which maximally utilise available clinico-pathological information and contextualise PCa within a patients context of competing risks. High quality models including treatment effect are overdue, and crucial if both undertreatment and overtreatment of prostate cancer is to be minimised.

Contributors DT conceived and designed the project under the supervision of VJG and PP. Screening and data extraction were performed by DT, SR and BB. Analyses were performed by DT, DT, SR and VJG wrote the manuscript. Each author reviewed and edited the final version.

Funding This work was supported through two research scholarships received through The Urology Foundation (DT & SR). The funders played no role in study design, collection, analysis, interpretation of data, writing of the report or in the decision to submit for publication. They do not endorse or accept any responsibility for the content.

Competing interests VJG was involved in developing the CPG model, but derives no monetary or other remuneration from the model, which is freely available and independently peer-reviewed.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Cancer Research UK. Prostate cancer statistics. <http://www.cancerresearchuk.org/cancer-info/cancerstats/types/prostate/>.
2. Bray F, Ferlay J, Soerjomataram I, et al. GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
3. Greenberg DC, Wright KA, Lophatanon A, et al. Changing presentation of prostate cancer in a UK population—10 year trends in prostate cancer risk profiles in the East of England. *Br J Cancer* 2013;109:2115–20.
4. Donovan JL, Hamdy FC, Lane JA, et al. Patient-Reported outcomes after monitoring, surgery, or radiotherapy for prostate cancer. *N Engl J Med* 2016;375:1425–37.
5. Greenberg DC, Lophatanon A, Wright KA, et al. Trends and outcome from radical therapy for primary non-metastatic prostate cancer in a UK population. *PLoS One* 2015;10:e0119494.
6. Huland H, Graefen M. Changing Trends in Surgical Management of Prostate Cancer: the end of overtreatment? *Eur Urol* 2015;68:175–8.
7. Lloyd T, Hounsborne L, Mehay A, et al. Lifetime risk of being diagnosed with, or dying from, prostate cancer by major ethnic group in England 2008–2010. *BMC Med* 2015;13:171.
8. D'Amico AV, Whittington R, Malkowicz SB, et al. Biochemical outcome after radical prostatectomy, external beam radiation

- therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* 1998;280:969–74.
9. Jhaveri FM, Zippe CD, Klein EA, *et al.* Biochemical failure does not predict overall survival after radical prostatectomy for localized prostate cancer: 10-year results. *Urology* 1999;54:884–90.
 10. Huang Y, Isharwal S, Haese A, *et al.* Prediction of patient-specific risk and percentile cohort risk of pathological stage outcome using continuous prostate-specific antigen measurement, clinical stage and biopsy Gleason score. *BJU Int* 2011;107:1562–9.
 11. Cooperberg MR, Pasta DJ, Elkin EP, *et al.* The University of California, San Francisco Cancer of the Prostate Risk Assessment score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy. *J Urol* 2005;173:1938–42.
 12. Stephenson AJ, Scardino PT, Eastham JA, *et al.* Preoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. *J Natl Cancer Inst* 2006;98:715–7.
 13. Kattan MW, Hess KR, Amin MB, *et al.* American Joint Committee on Cancer acceptance criteria for inclusion of risk models for individualized prognosis in the practice of precision medicine. *CA Cancer J Clin* 2016;66:370–4.
 14. Wishart GC, Azzato EM, Greenberg DC, *et al.* PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res* 2010;12:R1.
 15. Liberati A, Altman DG, Tetzlaff J, *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009;151:W–94.
 16. Moons KG, de Groot JA, Bouwmeester W, *et al.* Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014;11:e1001744.
 17. Wong SS, Wilczynski NL, Haynes RB, *et al.* Developing optimal search strategies for detecting sound clinical prediction studies in MEDLINE. *AMIA Annu Symp Proc* 2003;728–32.
 18. Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. *J Am Med Inform Assoc* 2001;8:391–7.
 19. Wolff RF, Moons KGM, Riley RD, *et al.* PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019;170:51–8.
 20. Moons KGM, Wolff RF, Riley RD, *et al.* PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med* 2019;170:W1–33.
 21. Margel D, Urbach DR, Lipscombe LL, *et al.* Is pathology necessary to predict mortality among men with prostate-cancer? *BMC Med Inform Decis Mak* 2014;14:14.
 22. Nguyen PL, Chen MH, Catalona WJ, *et al.* Predicting prostate cancer mortality among men with intermediate to high-risk disease and multiple unfavorable risk factors. *Int J Radiat Oncol Biol Phys* 2009;73:659–64.
 23. Feuer EJ, Lee M, Mariotto AB, *et al.* The Cancer Survival Query System: making survival estimates from the Surveillance, Epidemiology, and End Results program more timely and relevant for recently diagnosed patients. *Cancer* 2012;118:5652–62.
 24. Kutikov A, Cooperberg MR, Pacione AT, *et al.* Evaluating prostate cancer mortality and competing risks of death in patients with localized prostate cancer using a comprehensive nomogram. *Prostate Cancer Prostatic Dis* 2012;15:374–9.
 25. Cowen ME, Halasyamani LK, Kattan MW. Predicting life expectancy in men with clinically localized prostate cancer. *J Urol* 2006;175:99–103.
 26. Kerkmeijer LG, Monninkhof EM, van Oort IM, *et al.* PREDICT: model for prediction of survival in localized prostate cancer. *World J Urol* 2016;34:789–95.
 27. Gnanapragasam VJ, Lophatananon A, Wright KA, *et al.* Improving clinical risk stratification at diagnosis in primary prostate cancer: a prognostic modelling study. *PLoS Med* 2016;13:e1002063.
 28. Tewari A, Johnson CC, Divine G, *et al.* Long-term survival probability in men with clinically localized prostate cancer: a case-control, propensity modeling study stratified by race, age, treatment and comorbidities. *J Urol* 2004;171:1513–9.
 29. Howlader N, Mariotto AB, Woloshin S, *et al.* Providing clinicians and patients with actual prognosis: cancer in the context of competing causes of death. *J Natl Cancer Inst Monogr* 2014;2014:255–64.
 30. Feuer EJ, Rabin BA, Zou Z, *et al.* The Surveillance, Epidemiology, and End Results Cancer Survival Calculator SEER*CS: validation in a managed care setting. *J Natl Cancer Inst Monogr* 2014;2014:265–74.
 31. Gnanapragasam VJ, Bratt O, Muir K, *et al.* The Cambridge Prognostic Groups for improved prediction of disease mortality at diagnosis in primary non-metastatic prostate cancer: a validation study. *BMC Med* 2018;16:31.
 32. Cooperberg MR, Broering JM, Carroll PR. Risk assessment for prostate cancer metastasis and mortality at the time of diagnosis. *J Natl Cancer Inst* 2009;101:878–87.
 33. Shariat SF, Kattan MW, Vickers AJ, *et al.* Critical review of prostate cancer predictive tools. *Future Oncol* 2009;5:1555–84.
 34. Shariat SF, Karakiewicz PI, Roehrborn CG, *et al.* An updated catalog of prostate cancer predictive tools. *Cancer* 2008;113:3075–99.
 35. Green DA, Osterberg EC, Xylinas E, *et al.* Predictive tools for prostate cancer staging, treatment response and outcomes. *Arch Esp Urol* 2012;65:787–807.
 36. Lughezzani G, Briganti A, Karakiewicz PI, *et al.* Predictive and prognostic models in radical prostatectomy candidates: a critical analysis of the literature. *Eur Urol* 2010;58:687–700.
 37. Raymond E, O'Callaghan ME, Campbell J, *et al.* An appraisal of analytical tools used in predicting clinical outcomes following radiation therapy treatment of men with prostate cancer: a systematic review. *Radiat Oncol* 2017;12:56.
 38. Walz J, Gallina A, Bocciardi AM, *et al.* Nomogram predicting 10 year life expectancy after radical prostatectomy or radiation therapy for prostate cancer. *Journal of Urology* 2007;177:131–2.
 39. Albertsen PC, Fryback DG, Storer BE, *et al.* The impact of co-morbidity on life expectancy among men with localized prostate cancer. *J Urol* 1996;156:127–32.
 40. Fontanella P, Benecchi L, Grasso A, *et al.* Decision-making tools in prostate cancer: from risk grouping to nomograms. *Minerva Urol Nefrol* 2017;69:556–66.
 41. Kent M, Vickers AJ. A systematic literature review of life expectancy prediction tools for patients with localized prostate cancer. *J Urol* 2015;193:1938–42.
 42. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016;35:214–26.
 43. Kasivisvanathan V, Rannikko AS, Borghi M, *et al.* MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *N Engl J Med* 2018;378:1767–77.
 44. Hamdy FC, Donovan JL, Lane JA, *et al.* 10-Year Outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. *N Engl J Med* 2016;375:1415–24.
 45. Berry SD, Ngo L, Samelson EJ, *et al.* Competing risk of death: an important consideration in studies of older adults. *J Am Geriatr Soc* 2010;58:783–7.
 46. Grogan J, Gupta R, Mahon KL, *et al.* Predictive value of the 2014 International Society of Urological Pathology grading system for prostate cancer in patients undergoing radical prostatectomy with long-term follow-up. *BJU Int* 2017;120:651–8.
 47. Bostwick DG, Foster CS. Predictive factors in prostate cancer: current concepts from the 1999 College of American Pathologists Conference on Solid Tumor Prognostic Factors and the 1999 World Health Organization Second International Consultation on Prostate Cancer. *Semin Urol Oncol* 1999;17:222–72.
 48. Partin AW, Steinberg GD, Pitcock RV, *et al.* Use of nuclear morphometry, gleason histologic scoring, clinical stage, and age to predict disease-free survival among patients with prostate cancer. *Cancer* 1992;70:161–8.
 49. Epstein JI, Zelefsky MJ, Sjoberg DD, *et al.* A Contemporary prostate cancer grading system: a validated alternative to the Gleason Score. *Eur Urol* 2016;69:428–35.
 50. Moons KG, Altman DG, Vergouwe Y, *et al.* Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009;338:b606.
 51. Lorent M, Maalmi H, Tessier P, *et al.* Meta-analysis of predictive models to assess the clinical validity and utility for patient-centered medical decision making: application to the CAnCER of the Prostate Risk Assessment (CAPRA). *BMC Med Inform Decis Mak* 2019;19:2.
 52. Falzarano SM, Ferro M, Bollito E, *et al.* Novel biomarkers and genomic tests in prostate cancer: a critical analysis. *Minerva Urol Nefrol* 2015;67:211–31.
 53. Herlemann A, Washington SL, Eapen RS, *et al.* Whom to treat: postdiagnostic risk assessment with Gleason Score, Risk Models, and Genomic Classifier. *Urol Clin North Am* 2017;44:547–55.